2021-01-14

Metadata for Research Data

Carruthers, Matthew

For Research Data

Matt Carruthers

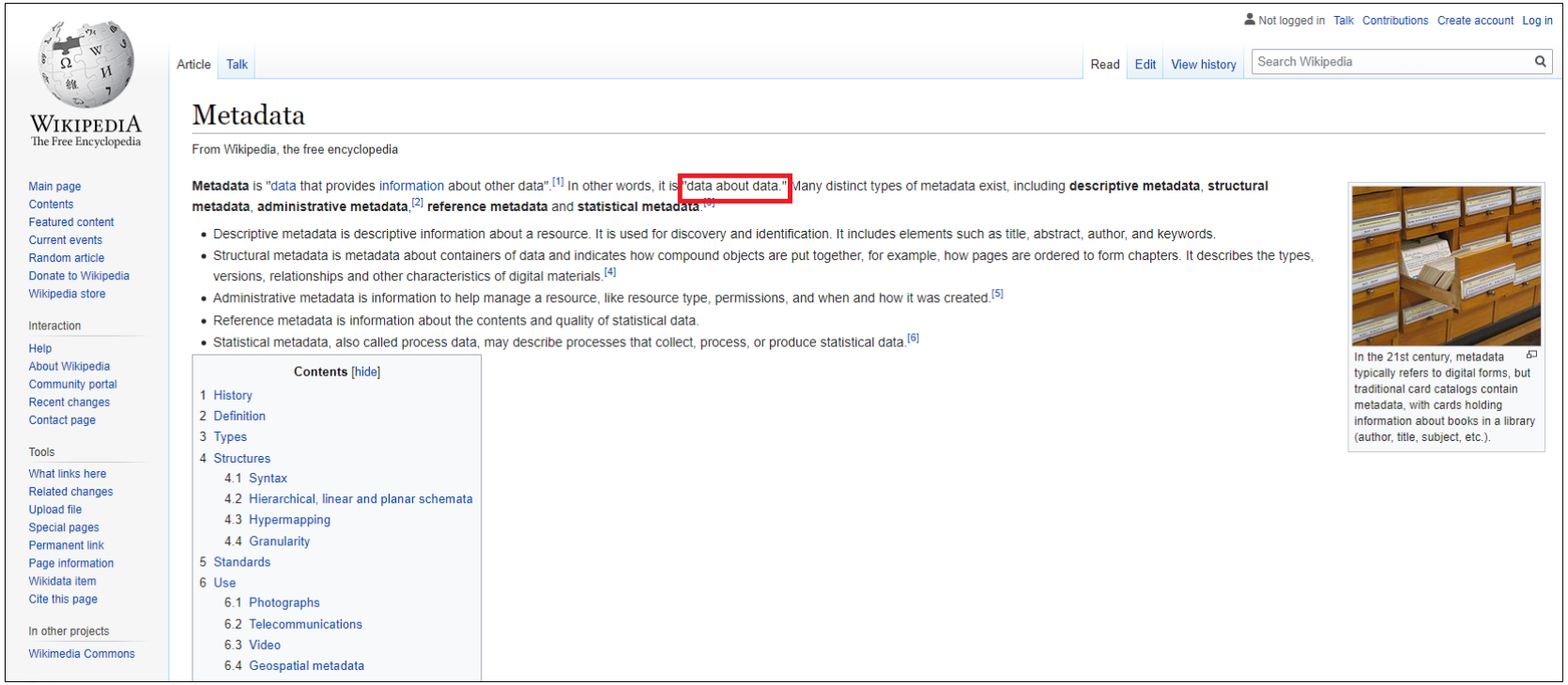
mcarruth@umich.edu

What we'll cover:

1. Defining metadata.
2. A brief history of the evolution of metadata.
3. Metadata in a digital environment.
4. Metadata and the research data lifecycle.
5. Identifying the right metadata standard.

What is metadata?

“Data about data.”



Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects
Wikimedia Commons

Metadata

From Wikipedia, the free encyclopedia

Metadata is "data that provides information about other data".[1] In other words, it is "data about data." Many distinct types of metadata exist, including **descriptive metadata**, **structural metadata**, **administrative metadata**,[2] **reference metadata** and **statistical metadata**.[3]

- Descriptive metadata is descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
- Structural metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.[4]
- Administrative metadata is information to help manage a resource, like resource type, permissions, and when and how it was created.[5]
- Reference metadata is information about the contents and quality of statistical data.
- Statistical metadata, also called process data, may describe processes that collect, process, or produce statistical data.[6]

Contents [hide]

1 History
2 Definition
3 Types
4 Structures
  4.1 Syntax
  4.2 Hierarchical, linear and planar schemata
  4.3 Hypermapping
  4.4 Granularity
5 Standards
6 Use
  6.1 Photographs
  6.2 Telecommunications
  6.3 Video
  6.4 Geospatial metadata

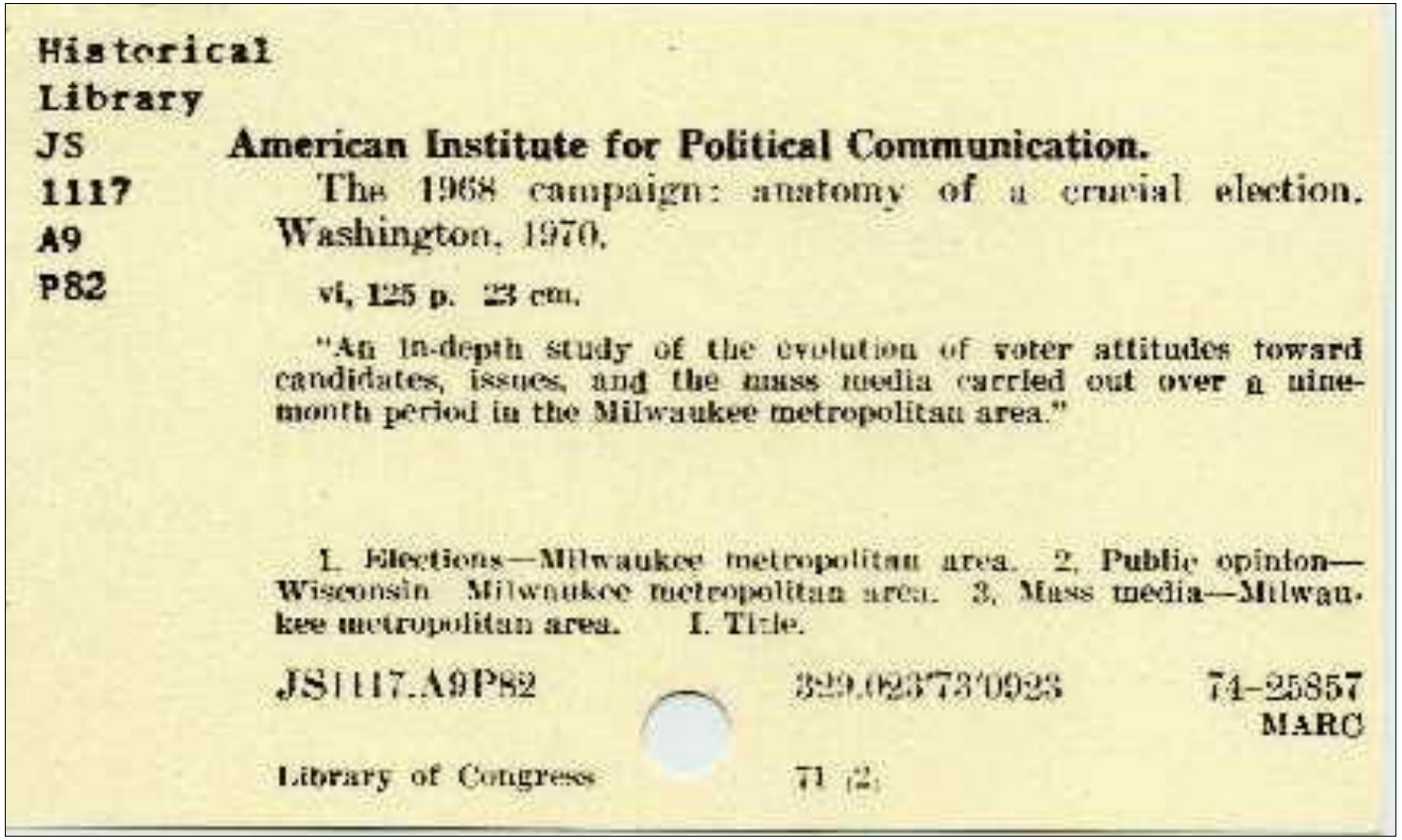
In the 21st century, metadata typically refers to digital forms, but traditional card catalogs contain metadata, with cards holding information about books in a library (author, title, subject, etc.).

What is metadata?

1. Metadata describes the content, quality, condition, and other characteristics of data.
2. Metadata is standardized, structured information about an object that facilitates functions associated with that object.

   (Discovery, management, rights and access control, preservation, reuse.)

Metadata in libraries: the past



Historical
Library
JS
1117
A9
P82

American Institute for Political Communication.

The 1968 campaign: anatomy of a crucial election. Washington, 1970.

vi, 125 p. 23 cm.

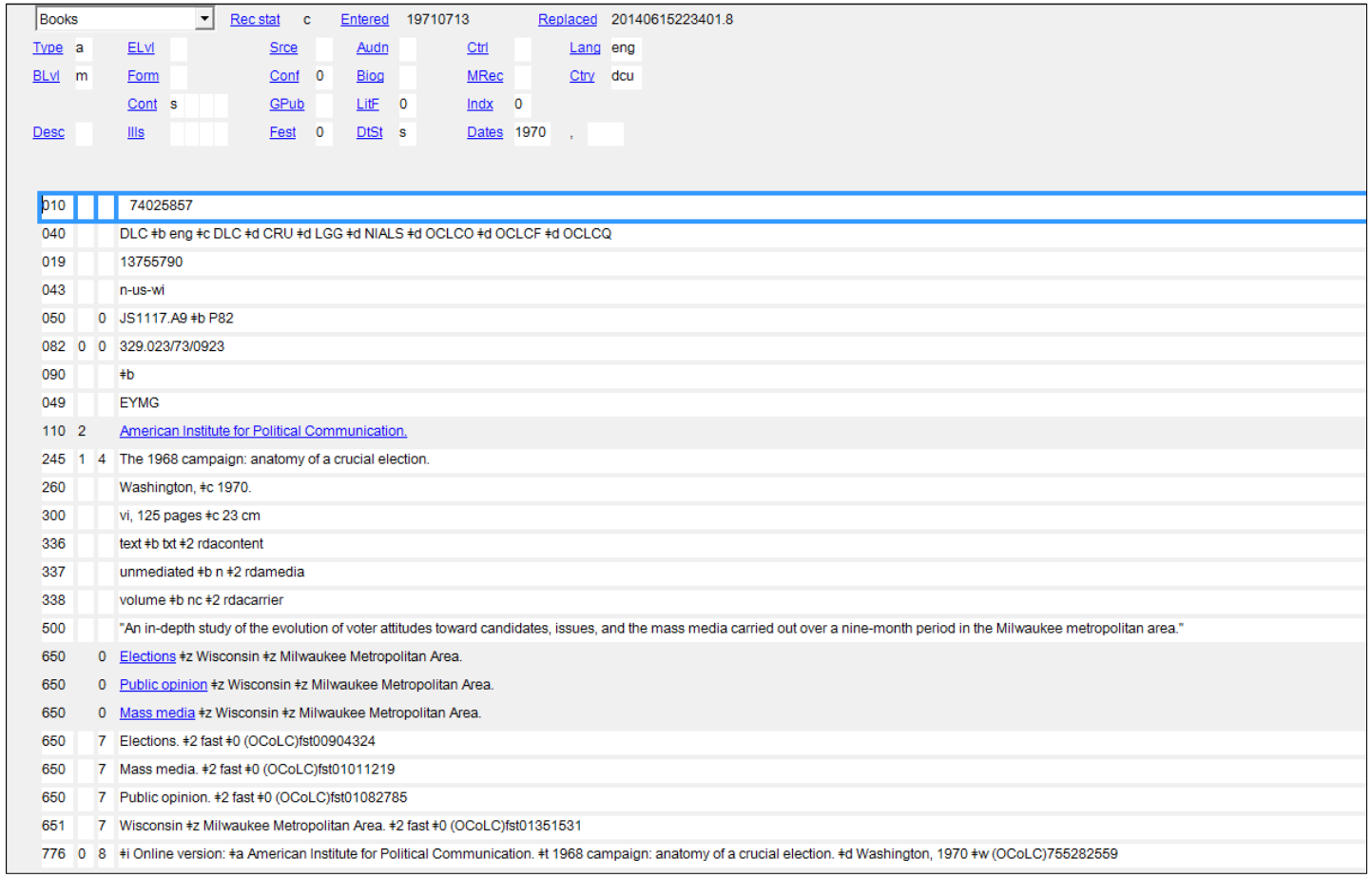
"An in-depth study of the evolution of voter attitudes toward candidates, issues, and the mass media carried out over a nine-month period in the Milwaukee metropolitan area."

1. Elections—Milwaukee metropolitan area. 2. Public opinion—Wisconsin—Milwaukee metropolitan area. 3. Mass media—Milwaukee metropolitan area. I. Title.

JS1117.A9P82 329.023730923 74-25857
MARC

Library of Congress 71 (2)

Metadata in libraries: the present



Books | Rec stat c | Entered 19710713 | Replaced 20140615223401.8

Type a | ELvl | Srce | Audn | Ctrl | Lang eng

BLvl m | Form | Conf 0 | Biog | MRec | Ctry dcu

Cont s | GPub | LitF 0 | Indx 0

Desc | Ills | Fest 0 | DtSt s | Dates 1970 ,

010			74025857
040			DLC ǂb eng ǂc DLC ǂd CRU ǂd LGG ǂd NIALS ǂd OCLCO ǂd OCLCF ǂd OCLCQ
019			13755790
043			n-us-wi
050		0	JS1117.A9 ǂb P82
082	0	0	329.023/73/0923
090			ǂb
049			EYMG
110	2		American Institute for Political Communication.
245	1	4	The 1968 campaign: anatomy of a crucial election.
260			Washington, ǂc 1970.
300			vi, 125 pages ǂc 23 cm
336			text ǂb txt ǂ2 rdacontent
337			unmediated ǂb n ǂ2 rdamedia
338			volume ǂb nc ǂ2 rdacarrier
500			"An in-depth study of the evolution of voter attitudes toward candidates, issues, and the mass media carried out over a nine-month period in the Milwaukee metropolitan area."
650		0	Elections ǂz Wisconsin ǂz Milwaukee Metropolitan Area.
650		0	Public opinion ǂz Wisconsin ǂz Milwaukee Metropolitan Area.
650		0	Mass media ǂz Wisconsin ǂz Milwaukee Metropolitan Area.
650		7	Elections. ǂ2 fast ǂ0 (OCoLC)fst00904324
650		7	Mass media. ǂ2 fast ǂ0 (OCoLC)fst01011219
650		7	Public opinion. ǂ2 fast ǂ0 (OCoLC)fst01082785
651		7	Wisconsin ǂz Milwaukee Metropolitan Area. ǂ2 fast ǂ0 (OCoLC)fst01351531
776	0	8	ǂi Online version: ǂa American Institute for Political Communication. ǂt 1968 campaign: anatomy of a crucial election. ǂd Washington, 1970 ǂw (OCoLC)755282559

Metadata in libraries: the present

New things to consider with digital objects and collections:

- How do you represent context in a digital environment?
- How do you facilitate long-term preservation?
- How do you track changes to digital objects over time?
  - Format migration
  - Versioning of files

Metadata in libraries: the present

Describing collections of objects rather than single objects/resources

- Archival collections
- Digital collections
- Digital exhibits

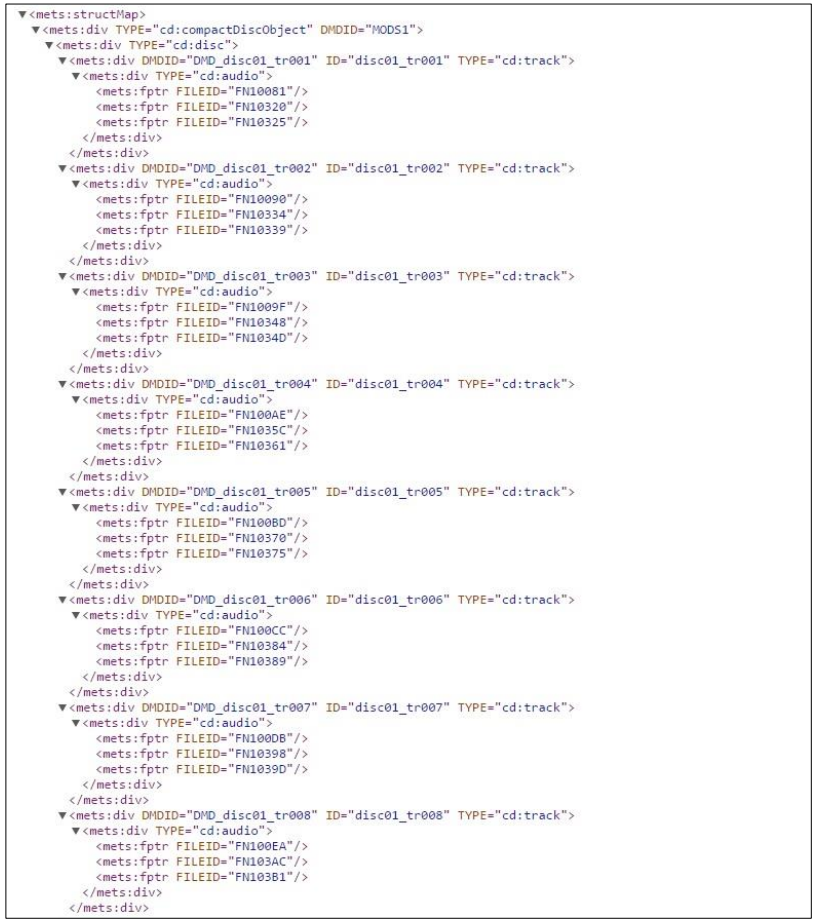
Encoded Archival Description (EAD)

Text Encoding Initiative (TEI)

Dublin Core (DC)

Metadata Object Description Schema (MODS)

Metadata Encoding & Transmission Standard (METS)



Types of Metadata

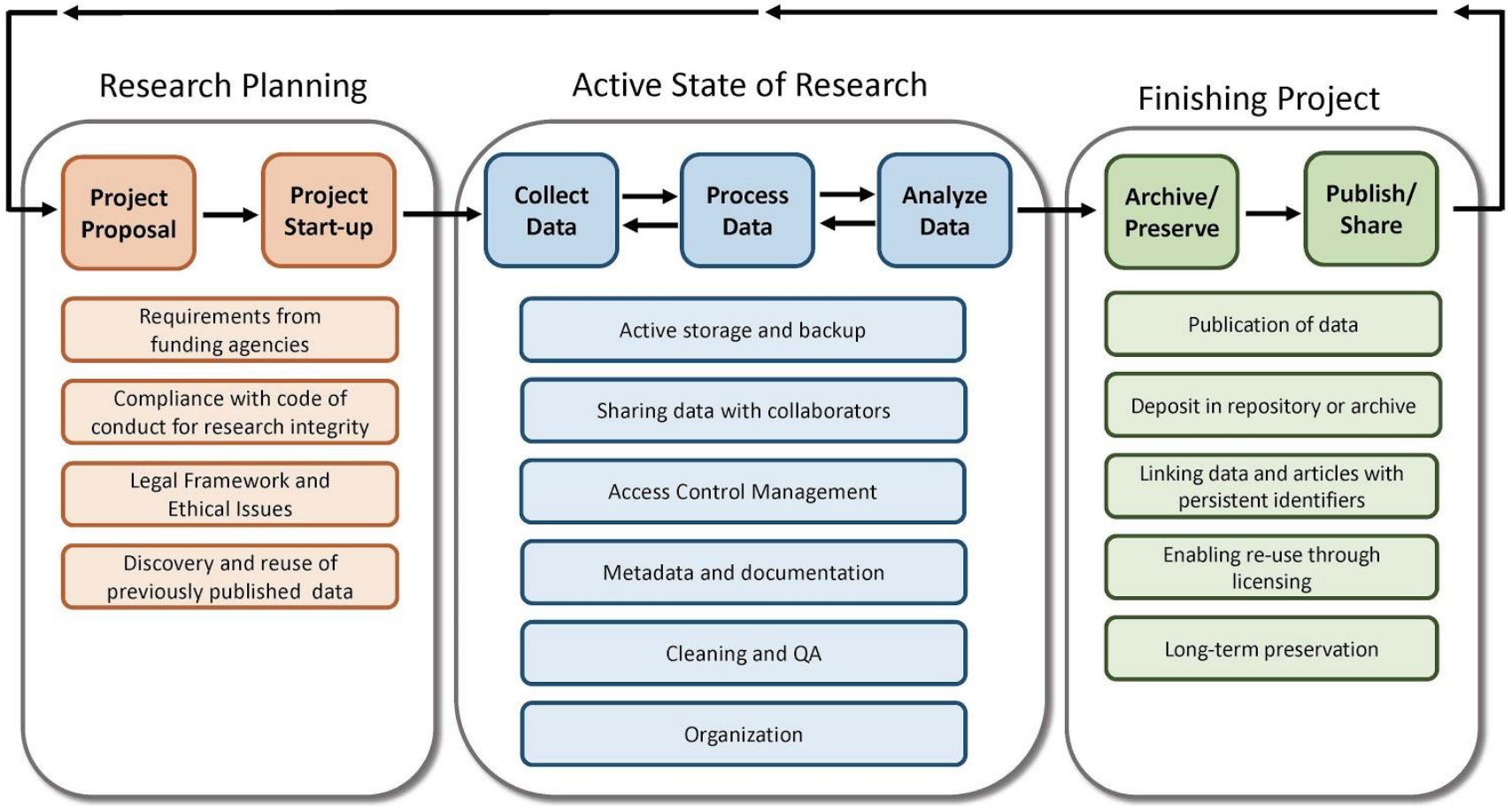
- Descriptive metadata
  - For finding or understanding a resource
- Administrative metadata
  - For long-term management of files
  - For decoding and rendering files
  - For documenting use and access rights
- Structural metadata
  - For documenting the relationships of parts of resources to one another

Types of Metadata

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability

Riley, Jenn. Understanding Metadata: What is Metadata, and What is it For: A Primer. NISO, 2017

Metadata and the Research Data Lifecycle



Adapted from Hüser, Falco Jonas; Elbæk, Mikael K.; Martinez Iavanchy, Paula (2016): DTU Research Data Life Cycle. figshare. Figure. https://doi.org/10.6084/m9.figshare.4258019.v1

Metadata and the Research Data Lifecycle

Levels of Metadata and Documentation:

1. **Study-level**: provides an overview of the research context and design, data collection methods, data preparation and results or findings.
2. **Data-level**: provides labeling and documentation of individual items, such as names and descriptions of variables, and explanations of codes and classification schemes used. It can be embedded within a data collection or recorded in an accompanying document.

Metadata and the Research Data Lifecycle

Metadata that can facilitate reuse will include information on:

- What research data exists
- Where it can be found
- How, when, and why the data was created
- Who created the data
- How to access the data
- What individual data points represent

Metadata and the Research Data Lifecycle

The difference metadata can make for comprehension and reuse:

Example 1

Example 2

Which example is easier to understand?

Which do you think has enough metadata to allow for reuse by other researchers?

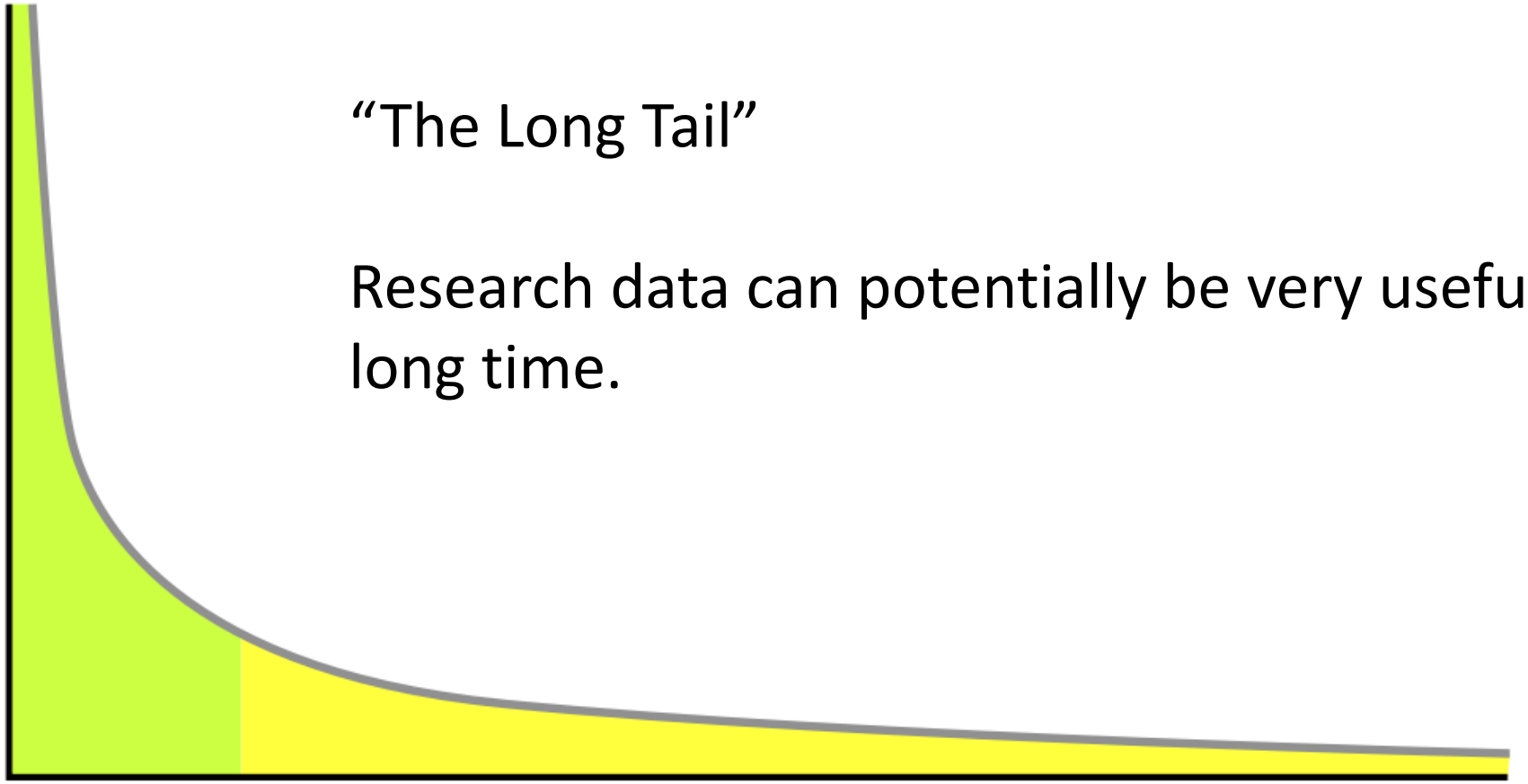
Why should researchers care?

Why should researchers care?

Scholarly Communication:

- Metadata can help fight the “Digital Data Deluge”.
- Make it discoverable or it might be lost forever.

Why should researchers care?



“The Long Tail”

Research data can potentially be very useful for a long time.

Why should researchers care?

The practical reasons:

- Potential for increase in data citations
- Funding agency requirements
- Can aid in the organization of your research data while you are creating/analyzing it

Identifying the right metadata standard

Too much to cover

ABCD - Access to Biological Collection Data
AgMES - Agricultural Metadata Element Set
AVM - Astronomy Visualization Metadata
CF (Climate and Forecast) Metadata Conventions
CIF - Crystallographic Information Framework
CIM - Common Information Model
CSMD-CCLRC Core Scientific Metadata Model
Darwin Core
DataCite Metadata Schema
DCAT - Data Catalog Vocabulary
DDI - Data Documentation Initiative
DIF - Directory Interchange Format
Dublin Core
EML - Ecological Metadata Language
FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata
FITS - Flexible Image Transport System
Genome Metadata
IVOA Photometry Data Model (PhotDM)
Simulation Data Model (SimDM)
Space-Time Coordinate (STC) Metadata for the Virtual Observatory
Astronomical Dataset Characterization Data Model (CharDM)
Simple Spectral Lines Data Model
IVOA Spectral Data Model
Observation Data Model Core Components (ObsCoreDM)
ISA-Tab
ISO 19115
MIBBI - Minimum Information for Biological and Biomedical Investigations
MIDAS-Heritage
OAI-ORE - Open Archives Initiative Object Reuse and Exchange
Observ-OM
Observations and Measurements
OME-XML - Open Microscopy Environment XML
PDBx/mmCIF – Protein Data Bank Exchange Dictionary and the Macromolecular
Crystallographic Information Framework
Protocol Data Element Definitions
PROV
QuDEx - Qualitative Data Exchange Format
RDF Data Cube Vocabulary
Repository-Developed Metadata Schemas
SDMX - Statistical Data and Metadata Exchange
SPASE Data Model

Identifying the right metadata standard

Questions to consider:

1. Who is the intended audience?
2. What are the research norms of the discipline?
3. Is there already an established standard for that discipline?
4. What legal or ethical requirements?

Identifying the right metadata standard

Resources to help you identify appropriate metadata standards:

Research Data Alliance Metadata Directory

Linked Open Vocabularies

Open Metadata Registry

FAIRsharing

BioPortal

Further reading

Research Data Curation Bibliography

Understanding Metadata: What is Metadata, and What is it For?: A Primer

Dublin Core Metadata Initiative Metadata Basics

DataONE Metadata Best Practices

Miller, Steven J.,. *Metadata for Digital Collections : A How-to-Do-It Manual*. Neal-Schuman Publishers, 2011.

Questions?